DES 0523-01 Information Design I: Data Visualization Fall 2022

1.1 Prepare Datasets

Download a **PDF of this page**.

Download a **step-by-step PDF** to complete this assignment. The PDF also includes links to videos. Below is some important software information and the list of files to upload.

Project 1 is a 3-week project composed of three parts:

This page is Part 1: 1.1 Prepare datasets
2 clean CVS (comma-separated value) files
3 plots (PDF)
1 R file)

Part 2: 1.2 Generate graphs (line, bar, scatterplot)
Part 3: 1.3 Clean up and format final graphs in Illustrator

Note: for spreadsheets, you must use Excel, not Google Sheets, because by editing in a web browser, HTML can introduce "dirty" characters like "/" that can cause errors later in the visualizations. The clean dataset file should always be checked with a text-only editor. Go to the **Software links** page to download all the necessary programs, including **Excel**, **BBEdit** (Mac) or **Notepad++** (PC), and **R**.

Summary of files

The two clean dataset files will be named:
lastName_brecan_75_17.csv
lastName_brecan_wb_2019.csv

The three plots will be named:
lastName_7517_plot1.pdf
lastName_wb_19_plot2.pdf
lastName_wb_19_scatterplot3.pdf

The R script will be named:
lastName_brecan.R

Some notes on the plots and info for dowloading and installing R and RStudio.

The goal of this project is to visualize the disparity of breast cancer mortality between White females and Black females, and show that the rate is almost 50% higher for Black, despite them having a lower incidence rate (fewer cases per 100K). We will not look at causes in this assignment, but it turns out that while lack of health care plays a big role, genetics also does.

After creating the two dataset files, plot a matrix for each using R, showing every possible combination of pairs of variables (columns). There is a bit of coding, just a few short lines. Plot also a scatterplot of just the white and black death rates.

Note: although R (or any other program) won't be required as the only program to use in the class, this little exercise will show that it's very reliable to just get a base graph plotted, and therefore I highly recommend it. Illustrator can then be used to clean up the base graph.

Download and install R (R Project): **https://www.r-project.org**
From one of the servers: **https://cran.r-project.org/mirrors.html**
For example Iowa State University: **https://mirror.las.iastate.edu/CRAN/**

The R site looks very nerdy but just make sure you download the correct version for your system, Mac or PC. Once on the mirror site,

download the latest version for Mac: **R-4.2.1.pkg** or the latest version for Windows: **R-4.2.1-win.exe**.

After installing R, install also RStudio, which is a graphic interface that runs on top of R. Download the free RStudio Desktop version: **https://rstudio.com/products/rstudio/download/**

After you have installed both, start only RStudio (R itself will launch and run, not visible, in the background).

Go to the **step-by-step PDF** for file and plot instructions.
Upload 6 files by the deadline:

1. lastName_brecan_75_17.csv
2. lastName_brecan_wb_2019.csv

3. lastName_7517_plot1.pdf
4. lastName_wb_19_plot2.pdf
5. lastName_wb_19_scatterplot3.pdf

6. lastName_brecan.R

Submission status

Attempt number	
Submission status	
Grading status	
Due date	Tuesday, August 30, 2022, 9:00 AM
Time remaining	
Last modified	
Submission comments	

Add submission

You have not made a submission yet.

◀ Attendance

Jump to...

1.1 Prepare Datasets Summary (PDF) ▶